Journal of the Southeast Asian Linguistics Society
JSEALS Vol. 10.2 (2017): 90-114
ISSN: 1836-6821, DOI: http://hdl.handle.net/10524/52415
University of Hawai'i Press

DIFFERENCES IN LANGUAGE USED BY DECEIVERS AND TRUTH-TELLERS IN THAI ONLINE CHAT

Montarat RUNGRUANGTHUM
School of Liberal Arts, King Mongkut's University of Technology Thonburi
linguistmona@yahoo.com

Richard Watson TODD
School of Liberal Arts, King Mongkut's University of Technology Thonburi
irictodd@kmutt.ac.th

Abstract
Deception detection, especially in online communication, is a burgeoning area of research, but most previous studies have focused on English. Therefore, in this paper, we investigate the applicability of English deceptive features to Thai and also examine whether there are any Thai-specific features not found in English which are associated with Thai deception in online communication. 96 Thai-language chat dialogues were analyzed with results suggesting that features identified in English deception research cannot be transferred to Thai. Two Thai-specific indicators of deception were also identified. The results have implications for theories of deception detection and for the transfer of research findings between languages.

Keywords: Thai, language differences, deception detection
ISO 639-3 codes: tha

1 Introduction
Developments in linguistics are often founded on work conducted in English or, to a lesser extent, other European languages. The theories and methods derived from such work are then applied to other languages. For instance, most of the original work in corpus linguistics focused on English leading to, for example, theories of collocation (e.g. Sinclair 1991) which have since been applied to other languages. At the level of specific research focus, an example is Pennebaker, Mayne and Francis' (1997) research into the relationship between language use and physical health conducted on English which has since been applied to other languages such as Italian (e.g. Celli & Poesio 2014) and Hungarian (e.g. Hargitai, Naszódi, Kis, Nagy, Bóna & László 2007).

The heavy focus on English as the source of linguistic theories can be considered valuable since the great amount of research into English gives the theories strong foundations. However, the dominance of English in generating linguistic theory could be viewed as a facet of linguistic imperialism (cf. Philipson 1992) and may also mean that certain linguistic features are overlooked as they do not exist or are occluded in English. One burgeoning area of interest which is highly dependent on original work conducted in English is deception detection. The recent growth of online communication and concern with the impacts of deception in such communication have led to a massive increase in the number of studies attempting to link the presence or absence of certain linguistic features with deceptive communication, which typically refers to message knowingly transmitted by a sender to create a false belief or conclusion (Buller & Burgoon 1996). The majority of this research has been conducted on English, but the findings from English have been applied to other languages without adding any additional features that might be associated with deception from that language.

In this paper, focusing on linguistic analyses of deception, we intend, first, to examine the applicability of findings from English to a typologically and historically different language, namely, Thai, and, second, to investigate if an analysis of the features of Thai deceptive language can highlight aspects not previously

found in research on English. In addition, we hope that our findings may provide directions for further research and applications concerning deceptive language in Thai.

2 The language of deception

The language of deception has become an area of interest in linguistic research since 2003, when Burgoon, Blair, Qin and Nunamaker (2003) published their influential study identifying language differences and patterns between deceivers and truth-tellers in face-to-face communication. In reviewing related studies of deceptive language, we searched for "deception detection", "online communication" and "linguistic-based cues" from three academic databases: IEEE, Proquest Dissertations and Theses Global and Springer. We have found 421 research studies (348 articles, 73 dissertations and theses) in total from 2003 to the present investigating specific features associated with deceptive language in online communication. The majority of these studies concern English, while only seventeen studies involve deception in other languages such as Italian and Chinese. For those languages, some results confirmed the findings from research into English (e.g. reduced use of self-references in deceptive Italian, e.g. Spence, Villar and Arciuli 2012, and reduced word quantity in deceptive Chinese, e.g. Zhou and Sung 2008), but these studies did not report any features specific to Italian or Chinese that might be associated with deceptive language. To sum up, 96% of the studies in deception detection have been conducted on English rather than on other languages. This clearly presents a strong bias in favor of English in previous research into deceptive language.

The language of deception has been investigated from three main perspectives: Cognitive Load Theory, Leakage Theory, and Interpersonal Deception Theory (IDT), each of which views the relationship between language and deception in different ways.

(1) Cognitive Load Theory argues that deceivers use more cognitive load than truth-tellers since they need to control their physical and emotional changes that may be indicative of deception (Zuckerman, DePaulo & Rosenthal 1981). Deceivers' brains are preoccupied with making up false stories so they delay their responses, make errors, and repeat words or phrases more often than truth-tellers (Vrij, Mann, Fisher & Leal 2008). With low cognitive load, truth-tellers, in contrast, can produce complex sentences and give specific details.

(2) Leakage Theory hypothesizes that language displays the speakers' emotions and inner feelings (Ochs & Schiefflin 1989). When speakers are telling lies in face-to-face or in online communication, they unconsciously leak verbal (written and/or spoken language) and nonverbal (gestures, facial expressions, and eye-gazing) clues because of their anxiety or guilt (Ekman & Friesen 1969; 1974). Previous research into deception detection suggests that deceivers use negative emotion words (e.g. terrible, dirty, horrible, etc.) and produce short responses more frequently than truth-tellers since they are afraid of getting caught (Newman, Pennebaker, Berry & Richards 2003).

(3) IDT views deception as an interactive process between senders and receivers, meaning that deceivers need to manage information (the manipulation strategy of content messages), behavior (the strategy of suppressing any signals of telling lies to appear normal), and image (the strategy of maintaining credibility) (Burgoon 2009:551-552). Deceptive strategies associated with this include reducing the use of first-person self-references (I, me, my) compared with truth-tellers (Wiseman 2007) and maintaining the flow of the conversation, which results in higher participation than truth-tellers in online conversation (Zhou 2005).

In this paper, we do not favor a particular theory. However, since the theories predict different linguistic features associated with deception, examining the features associated with deception that are found in this study provides evidence suggesting that the theory predicting those features is more likely to hold in the context under investigation.

In order to investigate the language of deception, previous studies have compared two sets of data, one from deceivers and one from truth-tellers, in a specific situation. In many cases, the situations designed to create a need for research subjects to be deceptive involve artificial role-playing, such as choosing items for survival after a plane crash or lying about a mock crime. The inauthenticity of such situations casts doubt on the validity of the findings and their applicability to authentic deception (Sip, Roeostorff, McGregor, & Frith, 2008). In this paper, we intend to mirror real-world use, and therefore, we decided to focus on initial encounters in Thai online chat.

The language of deception has been examined using two main approaches: hand analysis and automated analysis. The former refers to the traditional linguistic analysis used in forensic contexts, such as police interrogation and witness statements and need to be done by hand. A well-known example of hand-analysis

approach is Statement Validity Analysis (see Köhnken 2004), which involves an investigator analysing linguistic cues in a subject's testimony based on the premise that people intending to deceive frequently use generalized statements and try to relate events vaguely, whereas truth-tellers tend to give specific details (Rabon 2003). Automated linguistic analysis to distinguish language patterns between deceivers and truth-tellers in online communication often has the goal of preventing online crimes and increasing the accuracy of lie-detection programs. For example, Linguistic Inquiry Word Count (LIWC) is a computerized text-analysis program which has been widely used in studies of online deception and categorizes frequent words which show, for example, that deceivers use less lexical diversity than truth-tellers in online communication (e.g. Amos 2008, Zhou 2005, Zhou & Zhang 2004). In this paper, we will combine hand and automated analyses, with most features initially investigated through automated counting but with some hand analyses of features identified as salient to provide depth.

3 Features associated with online deceptive language

Most previous research has investigated deception in one of three main contexts: face-to-face conversation (e.g. Burgoon et al. 2003), formal contexts like police interrogations (e.g. Vrij, Mann, Kristen & Fisher 2007) and online communication (e.g. Zhou, Twitchell, Qin, Burgoon & Nunamaker 2003). Our research context places emphasis on deception in online communication due to the rising threats of online crimes, including sexual harassment, phishing for information, identity fraud and hijacking networks (Johnson 1997). This situation poses a serious concern for global internet users and those in Thailand, where the number of internet users grew from 18,300,000 to 27,653,927 between 2009 and 2014 (Internet Information Research 2015). According to Thailand Technology Crime Suppression Division (2011), internet crimes, especially identity fraud, increased 32% between 2010 and 2011. Many cybercrime cases in Thailand involve online deception using chat programs and social networks.

There have been numerous recent studies of deceptive language using language production to differentiate deceivers and truth-tellers in online communication (e.g. Hancock, Curry, Goorha & Woodworth 2005). The reported features found in English associated with deception can be divided into two major categories. One is communication behavior such as participation and topic initiation. The other is linguistic features such as the use of self-reference. In this section, we briefly present some of the reported features found in English which have been previously associated with online deception.

3.1 Communication behaviors

Some deceptive indicators are features which remain the same irrespective of the language used. Four examples of such features are the following.

3.1.1 Participation examines the total number of turns taken by deceivers and truth-tellers. Some studies (e.g. Zhou 2005, Zhou & Zhang 2004) found that deceivers use a greater number of turns and participate more actively than truth-tellers because they take a dominant position when engaging in online communication.

3.1.2 Productivity refers to quantity of language production, such as total numbers of words, total numbers of letters, average numbers of words per turns, and total numbers of sentences. Most studies claim that deceivers produce a higher number of words than truth-tellers in online conversation (e.g., Hancock, Curry, Goorha & Woodworth 2008; Zhou, Burgoon, Nunamaker & Twitchell 2004a) to manage the flow of the online conversation, to create mutuality, and to reduce the chat partner's suspicion (Burgoon, Buller & Floyd 2001). This point is controversial since some studies (e.g. Strömwall, Granhag & Hartwig 2004; Zhou & Zhang 2008) have found that truth-tellers produce more words with longer sentences than deceivers so as to present realistic detailed content from their experiences while responding in online conversation.

3.1.3 Spontaneous correction can be identified as an immediate correction when speakers replace a message in the next turn of the conversation. Some studies (Zhou 2005; Zhou et al. 2004a; Zhou et al. 2003; Zhou & Zhang 2004) suggest that deceivers produce less spontaneous correction than truth-tellers because of their preoccupation with generating untruthful information, so they are less likely to make spontaneous changes. In contrast, truth-tellers with a lower cognitive load are more conscious of misspelled words or any typing errors and correct their own mistakes immediately in online conversation.

3.1.4 Topic-initiation means how often deceivers open new topics when talking with their chat partners. Previous studies in deceptive language (Adkins et al. 2004; Zhou 2005; Zhou et al. 2003) have found that deceivers change the topics more frequently than truth-tellers in online conversation to avoid giving specific

information. Truth-tellers, however, interact smoothly with the current topic, continuing the flow of the conversation in online communication.

3.2 Linguistic features

Some linguistic features have attributes associated with deception (Zhou et al. 2004b). Five indicators of online deception identified in previous studies are presented as examples.

3.2.1 Frequency of self-references may be linked to deception. Deceivers use fewer first-person self-references than truth-tellers in online communication (Burgoon et al. 2003; Newman et al. 2003). The use of first-person self-reference involves ownership of a statement, so deceivers try to dissociate themselves from the online conversation. Compared with those telling lies, truth-tellers produce more first-person pronouns when telling personal stories (Wiseman 2007; Hancock et al. 2005; 2008).

3.2.2 Negative emotional words and negators show the speaker's state of mind when engaging in online conversation. Messages produced by deceivers contain more negative emotional words (e.g. poor, bad, terrible, etc.) and negators (e.g. no or not) than truth-tellers in online conversation, perhaps to reflect the deceivers' guilt and worry in telling lies (Hauch, Masip, Blandon-Gitlin & Sporer 2012; Zhou 2005). Zhou, Burgoon, Twitchell, Qin and Nunamaker (2004b) also report that truth-tellers use positive emotional words (e.g. happy, pretty, good, etc.) more frequently, compared with those telling lies in online conversation.

3.2.3 Discourse markers refer to the use of sentence connectors (e.g. although, besides, however, unless, or, whereas, but, etc.), which display the use of cognitive resources. The process of telling lies requires a high cognitive load (Richards & Gross 1999(so most deceivers produce simple sentences rather than complex sentences when engaging in online deception (Hancock et al. 2008; Newman et al. 2003). Truth-tellers, in contrast, use discourse markers more frequently to link ideas (Strömwall, Grandhag & Hartwig 2004).

3.2.4 Some linguistic features such as modal verbs and hedges indicate the speaker's uncertainty. Deceivers are indirect when giving untruthful information and use vague language, so they express more uncertainty in their messages than truth-tellers in online communication (Twitchell, Nunamaker & Burgoon 2004; Zhou & Zhang 2008). Those telling the truth are more direct and use affirmative sentences more frequently when engaging in online conversation (Strömwall et al. 2004).

3.2.5 Lexical diversity and complexity involves a higher level of sentence complexity. Deceivers use shorter words and often produce more simple sentences than truth-tellers. They may repeat the same words due to the increased cognitive load resulting from the process of deception (Duran, Hall, McCarthy & McNamara 2010, Zhou et al. 2004a; Zhou & Zhang 2008). Truth-tellers, on the other hand, produce longer and more complex sentences (Burgoon et al. 2003; Strömwall et al. 2004). Thus, the readability scores of texts or messages produced by deceivers are higher than for truth-tellers in online conversation, indicating that deceivers' messages are easier to understand.

These reported features associated with deception have been found in previous research focusing on English. It is unclear whether English features can be used to detect deception in other languages, like Thai. Therefore, in this paper, we aim to investigate their applicability in identifying deception in Thai online chat since there has been little research into deceptive language in Asian contexts (Rubin 2014) and, so far as we are aware, none into Thai. In addition to examining the application of English features, we also intend to find out whether there are any distinctive features of Thai not found in English associated with deceptive language use.

4 Applying English Features to Thai

The English deceptive features have been mostly applied to western languages, such as Italian, while a few studies have been conducted in Asian languages, like Chinese. Generally, deceptive research in other languages (e.g. Fornaciari & Poesio 2011; Zhou & Sung 2008) is based on English research and has followed the deceptive features found in English without adding any language-specific features from other languages. To the best of our knowledge, there has been no research done on deception detection of Thai native speakers while engaging in online conversation. This section, therefore, briefly summarizes the similarities and differences between Thai and English features and the uniqueness of the Thai language that might be problematic in examining deception detection in Thai online chat.

There are some similarities between Thai and English: (1) both are alphabetic languages, (2) both are written in a linear sequence from left to right, (3) both have no morphological case on nouns or verbs, (4) the

basic word order is Subject + Verb + (Object), (5) the negative construction is the same in both language, that is in front of main verb or Negative + Verb, (6) both language have passive structures, (7) the sentence types are similar (affirmative, negative, interrogative and imperative), (8) both use adverbs to modify verbs and adjective to modify nouns, (9) self-references and pronoun use are used in both languages, and (10) English and Thai share some basic word classes (e.g. nouns, verbs, adjectives, adverbs, prepositions, and conjunctions) (Dryer & Haspelmath 2013; Iwasaki & Ingkaphirom 2005; Thammawan 2008). These aspects allow text analysis to be conducted on Thai to identify features of deceptive language in online communication.

Other Thai characteristics which are different from English and need to be considered when applying these features to Thai: (1) there is no space or symbol to identify word and sentence boundary in Thai so all chat messages needed to be segmented before further analysis (see 6.5 Data preparation); (2) there is a politeness distinction indicated by the use of gender particles and pronoun use in Thai; (3) Thai does not change word order in interrogative questions, but used question words and ending interrogative particles; (4) Thai does not use auxiliary verbs in interrogative and negative sentences; (5) some Thai verbs can be used as adjectives, also known as "attribute verbs"; (6) Thai verbs does not change to indicate singular or plural subjects, and tense; (7) there is no clear distinction between adjectives and adverbs, and some sets of adverbs can function as adjectives in Thai; and (8) the use of self-references can reflect the gender of speakers.

In terms of what features in English can be applied to detect deception in Thai online chat, some English features can be applied to Thai straightforwardly because the process of identifying features is very similar in English and Thai, for example, communication behaviors such as participation and spontaneous correction. Other linguistic features, such as negators (no and not) produced by deceivers and truth-tellers follow the same operationalization in both English and Thai since there is a specific word list for negation (มิ /mí/, ไม่ /mâj/ 'no, not'(, which is often placed in front of the main verb in a sentence (Smyth 2002). However, some features cannot be operationalized in Thai online chat, for example, the inconsistency of using definite and indefinite articles (e.g. a, an, and the), and changing verb tense in story-telling (Rabon 2003). From all of the deceptive features identified in previous research, six features can be directly applied to Thai, nineteen features can be applied to Thai with minor adaptations, and two features are specific to Thai, as summarized in Table 1.

5 Features Specific to Thai

In addition to English deceptive features, we also investigate whether there are any specific features in Thai, not found in English, associated with online deception. An example of a distinctive feature in Thai is particles, which often occur in the sentence-final position. These features, not found in deceptive research into English, are frequently used to express politeness and to show formality and social relationships, as well as to signal the speaker's intention (Cooke 1989; Iwasaki & Ingkaphriom 2005). The most common particles are gender particles (GP): ครับ /kʰráp/ by male speakers, as shown in Example (1), and คะ /kʰá/ or ค่ะ /kʰâ/ by female speakers, as can be seen in Example (2). In the context of Thai online chat, there are variants in pronunciation and spelling such as ครับ /kráp/, often shortened as คับ /kʰáp/ or lengthened as คร๊าบ /kʰrá:p/ to get a chat partner's attention.

(1) (Gender particle used by male speakers[1])

มีแฟนยังครับ

มี|	แฟน|	ยัง|	ครับ|
mi:	fɛ:n	jaŋ	kʰráp
Have	boyfriend	yet	GP

'Have you got a boyfriend yet?'

[1] In the examples, the first line gives the actual writing (in which words are not separated), the second line provides the same information separated into words, the third gives the phonetic transcription, the fourth gives the literal translation of each word, and the last shows the intended meaning.

)2((Gender particle used by female speakers)

อ๋อ ชอบแมวกับหมาค่ะ					
อ๋อ|	ชอบ|	แมว|	กับ|	หมา|	ค่ะ|
ʔɔ̋:	tɕʰɔ̂:p	mɛ:w	kàp	mǎ:	kʰâ
Oh!	like	cats	and	dogs	GP

'Oh!)I(like cats and dogs.'

A second example is self-reference in Thai since Thai allows a wide range of first-person reference depending on gender, relative status, and formality of the situation. In contrast to English, firstly, there is a variety of first-person pronouns in Thai. Secondly, second- and third-person pronouns can be used for self-reference. Thirdly, there are certain nouns which can be used for self-references, especially kinship terms (e.g. น้อง /nɔ́:ŋ/ 'younger brother or sister'(and title nouns (e.g. อาจารย์ /ʔa: tɕa:n/ 'instructor'(. Finally, they are often omitted in a sentence. Overall, there are various choices concerning how to self refer in Thai which are not applicable to English. These self-references also include honorifics, as shown in Example (3), in which the chatter used พี่ /pʰî:/ 'older brother or sister' referring to himself to make online conversation friendlier although they are not actually siblings.

)3((Kinship term)

พี่ชื่อไอซ์ครับ			
พี่|	ชื่อ|	ไอซ์|	ครับ|
pʰî:	tɕʰɯ̂:	ʔaj	kʰráp
Older sibling	name	Ice	GP

'My name is Ice.'

To identify likely Thai specific features associated with deception, we conducted a pilot study examining the language differences between deceivers and truth-tellers using Thai online communication. The results of this study showed two Thai-specific features (various types of self-reference and gender particles) might be indicators associated with deception in Thai online chat, and provided a guidance what potential features in Thai that can be used to identify deception in the main study (see 6.1 Pilot studies).

To sum up, this paper investigates the language differences between deceivers and truth-tellers in Thai online chat by analyzing the communication behaviors (e.g. participation and topic initiation) and linguistic features (e.g. negative emotion words) found in previous research into English deceptive language use. We also extend previous work by examining Thai-specific features to identify deception in online communication. Table 1 lists and describes the features which were investigated for deception detection in Thai online chat and provides an overview of how these features were adjusted and implemented.

The features investigated fall in the gathered Thai data into two types for analysis. Some features, such as spontaneous correction and demonstratives, simply involve counting the frequency of the feature. We view these features as categories, so Table 1 shows the category operationalization for these features. For other features, there are several sub-categories as codes that can be counted. For example, specific chat features include codes for emoticons, reductions, transliterated words and code switching. Table 1 shows sample code operationalization for these features.

Table 1: *Features analyzed in the Thai data*

Communication Behaviors	Author(s)	Adaptation (if necessary)	Category Operationalization	Code Operationalization	Thai Examples
1. Participation	Burgoon et al. (2003) Zhou et al. (2004a) Zhou & Zhang (2004) Zhou (2005)	–	Number of turns in each conversation	–	–
2. Total no. of words	Burgoon et al. (2003) Zhou & Zhang (2004) Zhou (2005) Hancock et al. (2008)	Word segmentation becomes an operationalization issue because there are no spaces indicating word boundaries in Thai.	Number of words produced after word segmentation (Aroonmanakun 2002)	–	–
3. Total no. of letters	Burgoon & Qin (2006)	–	Number of letters produced	–	–
4. Total no. of syllables	Burgoon et al. (2003)	–	Number of syllables after syllable segmentation (Aroonmanakun 2002)	–	–
5. Average words per turn	Burgoon et al. (2003)	(see 3. Total no. of words)	Average words per turn	–	–
6. Opening new topics	Twtichell & Nunamaker (2004) Adkins et al. (2004) Zhou (2005)	–	Number of new topics introduced	–	(see Example 5)

7. Greeting types	Rungruangthum & Watson Todd (2011)	–	–	Frequency of greeting types in Thai online chat (Panyametheekul 2005) • Greeting • Asking Question • Inviting • Others (Mixed)	• **Greetings** สวัสดี /sà wàt diː/ 'Hi, Hello' หวัดดี /wàt diː/ 'Hi, Hello' • **Inviting** มาคุยกันเถอะ /maː kʰuj kan tʰɤ̀ʔ/ 'Let's talk'
8. Spontaneous correction	Zhou & Zhang (2004) Zhou (2005)	–	Number of spontaneous corrections	–	(see Example 6)
Linguistic Features	**Authors**	**Adaptation (If necessary)**	**Category Operationalization**	**Code Operationalization**	**Thai Examples**
9. Repetition	Burgoon et al. (2003) Strömwall et al. (2004) Zhou et al. (2004a) Zhou & Zhang (2008)	In Thai, repetition can be through repeating words, through repeating letters or through using a symbol (ๆ).	–	Frequency of repetitions based on Iwasaki and Ingkaphirom (2005) • Using symbols • Duplicating final letters • Duplicating words/phrases	• **Using symbol** ฮ่าๆ /hâː hâː/ 'Haha' • **Duplicating final letter** อืมมมมมมมม /ʔɯːm/ 'ummmmm' • **Duplicating words/phrases** ไม่ไม่ /mâj mâj/ 'No No'
10. Specific chat features	Hoonchamlong (2005)	Some chat features derived from English can be directly applied to Thai, e.g., emoticons. But transliterated words are added into this category specific to Thai.	–	Frequency of specific chat features (Hoonchamlong 2005) • Emoticons • Reduction • Abbreviation • Swear words • Transliterated words • Code switching	• **Transliterated Words** Pood Thai dai 'can speak Thai' • **Code switching** Deceiver: Hi Naïve partner: พูดไทยได้ปะ /pʰûːt tʰaj dâj pàʔ/ 'speak Thai, can you?' Deceiver: โทษๆ ลืม /tʰôːt tʰôːt lɯːm/ 'Sorry, (I) forgot' The participant as deceiver switched from English to Thai.

11. Lexical complexity	Zhou et al. (2004a) Zhou et al. (2004b) Burgoon et al. (2003) Burgoon & Qin (2006) Duran et al. (2010)	The list of content words and functional words in Thai are based on Iwasaki and Ingkaphirom (2005).	–	Measures of lexical complexity • Total number of content words • Total number of functional words	–
12. Lexical diversity	Zhou et al. (2004a) Zhou et al. (2004b) Burgoon et al. (2003) Burgoon & Qin (2006)	(see 3. Total no. of words)	Type-token ratio	–	–
13. Sentence types	Strömwall et al. (2004) Twitchell et al (2004) Zhou (2005) Rungruagnthum & Watson Todd (2011)	Sentence identification is based on Iwasaki and Ingkaphirom (2005).	–	Frequency of sentence types based on Iwasaki and Ingkaphirom (2005) • Affirmative • Negative • Interrogative • Imperative	–
14. Sentence complexity	Burgoon et al. (2003) Zhou et al. (2004b) Burgoon & Qin (2006) Duran et al. (2010)	(See sentence types)	–	Frequency of sentence complexity based on Iwasaki and Ingkaphirom (2005) • Compound sentence • Complex sentence	–

15. Nouns	Burgoon et al. (2003) Zhou et al. (2004b) Burgoon & Qin (2006) Zhou & Zhang (2008)	–	–	Frequency of different types of nouns (Sornlertlamvanich, Takahashi & Isahara 1999) • Proper noun • Cardinal number noun • Ordinal number noun • Label noun • Common noun	–
16. Verbs	Adkins et al. (2004) Zhou et al. (2004a) Zhou et al. (2004b) Burgoon & Qin (2006) Hancock et al. (2008)	Some Thai verbs can be used as attribute verbs.	–	Frequency of verbs (Sornlertlamvanich, Charoenporn & Isahara 1997) • Action Verbs • Stative Verbs	• **Action Verbs** เรียน /riːan/ ‘learn’ ไป /paj/ ‘go’ • **Stative Verbs** เป็น /pen/, อยู่ /jùː/ ‘be’ ได้ยิน /dâj jin/ ‘hear’
17. Demonstratives	Strömwall et al. (2004) Rungruangthum & Watson Todd (2011)	–	Number of demonstratives	–	• **Demonstrative** นี่ /nîː/, นี้ /níː/ ‘This, These’ นั่น /nân/, นั้น /nán/ ‘That, Those’
18. Question words	Twitchell et al. (2004) Rungruangthum & Watson Todd (2011)	Some ending particles can also be used to interrogative questions.	–	Frequency of question words based on Iwasaki and Ingkaphirom (2005) • Ending interrogative questions • Wh-questions	• **Ending interrogative questions** ไหม /măj/, มั้ย /máj/ ‘Yes/No question’ เหรอ /rɤ̌ː/ ‘Yes/No question’ • **Wh-Questions** ใคร /kʰraj/ ‘who’ ไหน /năj/ ‘which’ อะไร /ʔàʔ raj/ ‘what’ กี่ /kìː/ ‘how many’

19. Question types	Twitchell et al. (2004) Rungruangthum & Watson Todd (2011)	(see 18. Question words)	–	Frequency of questions types based on the categorization of Smyth (2002) • Wh-questions • Yes/ No questions • Alternative questions • Omission	• **Yes/No question in Thai** ไป ด้วย ไหม /paj dûaj măj/ Go too do you? 'Do you want to go too?'
20. Expressiveness	Zhou et al. (2004b)	There is no morphological identification between adjectives modifying nouns and adverbs modifying verbs	–	Frequency of expressive words (Iwasaki & Ingkaphirom 2005) • Adjectives • Adverbs	–
21. Negator	Burgoon et al. (2003) Strömwall et al. (2004) Twitchell et al. (2004) Zhou (2005) Zhou & Zhang (2008) Hancock et al. (2008)	–	Number of negative words	–	–
22. Frequency of overall pronoun use	Burgoon et al. (2004) Hancock et al. (2008) Newman et al. (2003) Wiseman (2007) Zhou (2005) Zhou et al. (2004).	–	–	Frequency of pronoun use based on Iwasaki and Ingkaphirom (2005) • 1st personal pronouns • 2nd personal pronouns • 3rd personal pronouns	• **2nd personal pronouns** เธอ /tʰɤ:/ 'you' (female speakers) แก /kɛ:/ 'you (male/female speakers) นาย /na:j/ 'you' (male speakers) • **3rd personal pronouns** เค้า /kʰá:w/, เขา /kʰăw/ 'he, she'

23. Self-reference frequency	Burgoon et al. (2004) Hancock et al. (2008) Newman et al. (2003) Wiseman (2007) Zhou (2005)	There are many ways in Thai to refer to first-person pronouns, which are related to interlocuters' gender and relationship.	Number of -person self-reference	–	• **First self-references** ฉัน /tɕʰǎn/ 'I' (female speaker) เรา /raw/ 'I' (male/female speaker) ผม /pʰǒm/ 'I' (for male speaker)
24. Modal Verbs	Twitchell et al. (2004) Zhou (2005) Burgoon & Qin (2006)	There are specific word sets that can be used as modal verbs in Thai, related to the level of confidence.	–	Frequency of modal verbs categorized into three levels of confidence based on Iwasaki and Ingkaphirom (2005) • High • Mid • Low	• **High confidence** ต้อง /tôŋ/ 'must' จำเป็น /tɕam pen/ 'necessary' • **Mid confidence** ควร /kʰuːan/, น่า /nâː/ 'should' • **Low confidence** อาจ /ʔàːt/, คง /kʰoŋ/ 'may'
25. Interjections	Burgoon et al. (2003) Hancock et al. (2008) Strömwall et al. (2004)	–	–	Frequency of Thai interjections based on the study of Yingnam, Premchaisawadi, and Kreesuradej (2009) • Positive • Neutral • Negative	• **Positive** ว้าว /wáːw/ 'Wow!' เย้ /jéʔ/ 'Yeah!' • **Neutral** อ๋อ /wɤ̌ːj/ 'Oh!' อือ /ʔɯː/ 'Right!' • **Negative** โอ๊ย /ʔôːj/ 'Ouch!' ว้าย /wáːj/ 'Oh no! '

Thai specific features	**Authors**	**Adaptation (If necessary)**	**Category Operationalization**	**Code Operationalization**	**Thai Examples**
26. Self-reference types	Rungruangthum & Watson Todd (2011)	Thai specific feature	–	Frequency of types of self-references based on Iwasaki and Ingkapirom (2005) • 1st personal pronouns • Kinship terms • Title name • Using Nickname • Ellipsis pronoun	• **Kinship Terms** พี่ /pʰîː/ ‘Older sibling’ น้อง /nɔ́ːŋ/ ‘Younger sibling’ • **Title Name** อาจารย์ /ʔaː tɕaːn/, ครู /kʰruː/ ‘teacher’ • **Nickname** น้ำตาล /nám taːn/ ‘Female nickname ชัย /tɕʰaj/ ‘Male nickname’
27. Particles	Rungruangthum & Watson Todd (2011)	Thai specific feature	–	Frequency of particles based on Iwasaki and Ingkapirom (2005) • Gender particles • Information-oriented particles • General-oriented particles • Action- oriented particles	• **Gender particles** คะ /kʰáʔ/, จ้า /tɕâː/ ‘female speaker’ ครับ /kʰráp/, ฮะ /háʔ/ ‘male speaker’ • **Information oriented particle** นะ /náʔ/**,** สิ /sìʔ/ ‘persuasive’ • **General-oriented particles** หรอก /rɔ̀ːk/ ‘counterargument’ ละ /láʔ/ ‘conclusion’ • **Action-oriented particles** เถอะ /tʰɤ̀ʔ/, เหอะ /hɤ̀ʔ/ ‘suggestion’

6 Research Methodology

6.1 Pilot Studies

In this research, pilot studies were conducted to investigate linguistic features associated with deception in Thai online chat and to identify the problematic issues while developing an experimental and data collection procedure.

The first pilot study was originally done in 2010 with the aim of identifying potential language differences between deceivers and truth-tellers in Thai on instant messaging. The voluntary participants were five native speakers of Thai, studying at the university level. They engaged in two chat sessions with another random chat partner on MSN Messenger using pseudonyms as deceivers and truth-tellers. All chat messages were then compared and analysed for similarities and differences in language use between deceivers and truth-tellers in Thai online chat. The results of the pilot study showed that truth-tellers used verbs, adverbs and interjections more frequently than deceivers in Thai online chat. In addition, there were Thai-specific features, namely, gender particles and ways of expressing self-reference associated with Thai online deception, indicating that these may be potential features to investigate in the main study (Rungruangthum & Watson Todd 2011).

Further pilot studies were conducted to consider issues of context, ethics and task instructions (see 6.3 Task and Setting).

6.2 Participants

Ninety-six Thai native speaker university students participated in this study. Their ages ranged from 18 to 23 years old, and they are representative of the age demographic of the most frequent internet users in Thailand. The National Electronics and Computer Technology Center, Thailand (2010) reports that 45% of Thai internet users fall into this age group. They were divided into two groups: 48 chatters and 48 naïve partners. In this study, the chatters were participants taking a role of either truth-tellers or deceivers convincing their chat partners that they were someone else while engaging in Thai online chat. The naïve partners were participants who were assigned as chat partners taking part in the online conversation without knowing the true objective of the study.

These two groups of participants were divided into twenty-four sub groups. Within each sub group, there were two chatters and two naïve partners randomly paired to engage in two chat sessions on MSN Messenger with the pairings changing between chat sessions. All the participants did not know who their partners were before chatting. In order to simulate the instant messaging environment, participants in each sub-group were assigned to separate rooms where they could not see each other. None of participants were aware of the roles of the other member, and all of them were given a detailed description of the task. The data collection was conducted in a controlled setting in which chatters switched their roles as truth-tellers or deceivers while chatting with naïve partners.

For participants as chatters, they were informed of the objective of the study, namely, to identify differences in language use between deceivers and truth-tellers in Thai online chat. In one chat session, they were assigned to be truth-tellers providing only truthful information, while in the other chat session, they were assigned to be deceivers convincing naïve partners that they were someone else in online communication. Naïve partners, in turn, were informed that they were recruited for a study on the use of Thai language in online communication and asked to chat with each partner for ten minutes. All participants were informed that their chat messages would be recorded before they signed a consent form for participating in the experiment.

6.3 Task and Setting

To investigate differences in language use between deceivers and truth-tellers in Thai online chat, we needed a task that mirrors real world online communication and therefore decided to use a typical first encounter. Factors including ethics, chat program, and task design had to be identified before the data collection of the study. One of our primary concerns is ethical considerations in deceptive research, in particular, whether participants would be offended by being asked to tell lies while engaging in an experimental study of deception detection in Thai online chat. From our pilot studies (Rungruangthum, Watson Todd &

Aroonmanakun 2011), the interview data showed that participants as deceivers were not worried about telling lies because they used pseudonyms, and lying seemed to be an ordinary phenomenon in online conversation. Participants as naïve partners also added that they were willing to participate in this kind of the study after the purpose was explained to them.

To find out which chat program should be used in this study, we then conducted a survey of users' preferences. Of five options given, it was found that 98 of 100 respondents used MSN Messenger as a tool in Thai online communication. In addition to its popularity and user' familiarity, it also allowed us to randomly pair up chatters with naïve partners and record chat dialogues for later analysis.

We also conducted pilot studies into how different types of instructions could affect the data collection. From the pilot study (Rungruangthum, Watson Todd & Aroonmanakun 2011), if participants as deceivers were instructed "to provide only untruthful information in online conversation" or "to pretend to be someone else in online conversation", they told lies based on fantasy stories which could be easily detected by the naïve partners, resulting in unreal communication. Instructions "to convince their chat partners that they were someone else in online communication" resulted in communication that was not obviously deceptive. This instruction then was used in this study. To confirm that deceivers in this study followed the instruction, we also checked through chat dialogues and found that all deceivers told lies by convincing their partners that they were someone else.

6.4 Data Collection

The data collection of deception detection was conducted in 2011 at a controlled computer laboratory in a Thai university to identify the language differences between deceivers and truth-tellers in Thai online chat. Participants as chatters were led to isolated computer terminals in the computer room where they would perform the task. The chatters were informed of the true objective of the study, namely, to examine the differences in language use between deceivers and truth-tellers in Thai online chat, so participants as chatters needed to switch their roles either as truth-tellers or deceivers (assigned to a random session) between the two 10-minute chat sessions. Participants as truth-tellers were instructed to provide only truthful information when having an online conversation. Participants as deceivers were instructed to convince their naïve partners that they were someone else in an online conversation. Before the data collection, the participants as truth-tellers and deceivers had five minutes to prepare themselves to provide only truthful or untruthful information

The other group members as naïve partners were led to another computer room where they were informed that they would participate in two chat sessions with unknown partners to investigate the use of Thai chat language on MSN Messenger. In one of the 10-minute chat sessions, naïve partners were paired up with either truth-tellers or deceivers and switched their chat partners in the other 10-minute chat session. The participants as naïve partners also had five minutes to read the instructions and ask if they had any questions before that data collection started.

All participants then logged into MSN Messenger by using given pseudonyms at their computer stations and engaged in a 10-minute chat session with their chat partners. Once participants had finished the first 10-minute chat session, all participants were asked to remain at their computer terminals before switching to another unknown chat partners and engaging in another 10-minute chat session.

The online interactions between chatters as either truth-tellers or deceivers paired with naïve partners were automatically recorded on the server. After all participants finished the task, the true objective of the study, that is, investigating the differences in language use between deceivers and truth-tellers in Thai online chat, was revealed and naïve partners were allowed to withdraw from the study if they wished. The whole experiment was completed in 30-35 minutes.

6.5 Data Preparation

The ninety-six chat dialogs obtained from Thai deceivers and truth-tellers were first converted into text files before analysis. Four corpora were created: full conversations involving deceivers, full conversation involving truth-tellers, deceivers' turns only, and truth-tellers' turns only. These files were segmented into words using a word segmentation program (Aroonmanakun 2002) because in Thai writing, there are no spaces put between words, and the words were tagged for parts of speech using Smart Word Analysis for Thai (Charoenpornsawat 2003) and Simple Concordance Program (Reed 2012). From this, the 96 dialogs contained 25,479 words, and 7,720 turns, giving 3.003 words per turn.

Regarding Thai data preparation for further analysis, there were orthographic challenges. Thai writing consists of a string of symbols in which there are no explicit word or sentence boundaries (Tesprasit, Charoenpornsawat & Sornlertlamvanich 2003). This causes difficulty in identifying words and sentences used by deceivers and truth-tellers in Thai online chat. Therefore, it is necessary to separate Thai statements into smaller units before conducting a word frequency count between deceivers and truth-tellers. In this paper, Thai chat messages were segmented using the Thai Word Segmentation program (Aroonmanakun 2002), as shown in Example (4). However, the problems with idiosyncrasies in chat, such as duplicating letters for effect and misspellings, affect this division, and so the segmentation procedure was also hand-checked.

)4((Segmented Words)

อ๋อ พี่อยู่ปี4แล้ว

อ๋อ|	พี่	อยู่|	ปี|	4|	แล้ว|
ʔǒ:	pʰîː	jûː	piː	sìː	lɛ́ːw
Oh	older sibling	be	year	four	already

'Oh! I am in the fourth year already.'

In this utterance, the participant as deceiver produced six words. The data after segmentation was used for the analysis.

7 Data Analysis

As mentioned earlier, the investigated features used by the Thai deceivers and truth-tellers in this study can be separated into two groups: features with no sub-categories and features with sub-categories. For features where there are no sub-categories (e.g. participation and demonstratives), percentage differences between deceivers and truth-tellers were calculated. We then set an arbitrary cut point of fifteen percentage points difference between these two groups for the features to be interpreted as showing a difference between deceivers and truth-tellers. For features with sub-categories (e.g. verbs and question words), a chi-square test was used to examine whether there are any statistically significant differences in the language used by deceivers and truth-tellers in Thai online chat. However, in some cases, the assumptions made by chi-square were not met, for example, when expected frequencies were less than five (Yates, Moore & McCabe 1999). In these cases, categories that were logically associated were grouped together. To avoid Type I errors (or false positives), a significance level of .01 was used in this study. Effect size was also calculated using phi-coefficient to provide an indication of the strength of the findings (Larson-Hall 2012). Following Cohen (1992), we interpreted an effect size of .10 as small, .30 as medium, and .50 as large.

8 Results

The differences in the use of the investigated features by Thai deceivers and truth-tellers in online chat are presented in Table 2. In this table, for those features with no sub-categories, the percentage differences between deceivers and truth-tellers are calculated with a difference of at least 15% taken as indicating a marked difference (shown with an asterisk). For those features with sub-categories, the chi-square value was calculated with a significance level of 0.01 (significant values are marked with an asterisk). Features showing a difference between deceivers and truth-tellers are illustrated through examples from the data set.

Table 2: *Investigated features used by deceivers and truth-tellers in Thai online chat.*

Differences in Communication Behaviors									
	Features	**Deceivers**	**Truth-tellers**	**Total**	***df***	**Chi Square**	***P*-Value**	**Effect Size**	**Percentage Difference**
1.	Participation	1987	1998	3985	–	–	–	–	-0.55
2.	Total No. of Words	6518	6378	12896	–	–	–	–	2.20
3.	Total No. of Letters	22961	20955	43916	–	–	–	–	9.57
4.	Total No. of Syllables	7533	7297	14830	–	–	–	–	3.23
5.	Average Word per Turn	3.28	3.18	3.23	–	–	–	–	3.14
6.	Open New Topics	573	495	1068	–	–	–	–	15.76*
7.	Greeting Types	48	48	96	1	0.27	.60	.08	–
8.	Spontaneous Correction	34	29	63	–	–	–	–	17.24*
Differences in Linguistics Features									
	Features	**Deceivers**	**Truth-tellers**	**Total**	***df***	**Chi Square**	***P*-Value**	**Effect Size**	**Percentage Difference**
9.	Repetition	470	571	1041	2	3.95	.13	.28	–
10.	Specific Chat Features	398	345	743	3	0.79	.85	.12	–
11.	Lexical Complexity	5966	5765	11731	1	0.12	.73	.05	–
12.	Lexical Diversity	0.18	0.20	0.15	–	–	–	–	-10.00
13.	Sentence Types	961	1009	1970	3	5.05	.17	.32	–
14.	Sentence Complexity	215	223	438	1	0.87	.35	.13	–
15.	Nouns	1027	923	1950	1	1.30	.25	.16	–
16.	Verbs	1007	1008	2015	1	0.27	.60	.07	–
17.	Demonstratives	92	77	169	–	–	–	–	19.48*
18.	Question Words	529	549	1078	1	2.53	.11	.22	–
19.	Question Types	360	391	751	3	1.40	.70	.17	–
20.	Expressiveness	583	585	1168	1	0.05	.82	.03	–
21.	Negators	179	181	360	–	–	–	–	-1.10
22.	Freq. of overall pronoun use	300	240	540	2	3.77	.15	.28	–
23.	Freq. of self-references	178	132	310	–	–	–	–	34.85*
24.	Modal Verbs.	93	110	203	2	11.92	.00*	.50	–
25.	Interjections	393	430	823	2	3.40	.18	.26	–
Differences in Thai Specific Features									
	Features	**Deceivers**	**Truth-tellers**	**Total**	***df***	**Chi Square**	***P*-Value**	**Effect Size**	**Percentage Difference**
26.	Self-reference types	845	802	1647	3	16.73	.00*	.59	–
27.	Particles	740	635	1375	3	37.53	.00*	.88	–

The first set of results concerns those features which have no sub-categories and therefore were analyzed by percentage differences in order to see the difference in the use of this group of features between deceivers and truth-tellers in Thai online chat. It can be seen from the data in Table 2 that there were noticeable differences between deceivers and truth-tellers for two communication behaviors. Thai participants as deceivers, in this study, changed topics (Example 5) and made spontaneous correction (Example 6) in online conversation more than truth-tellers. In Example 5, after two turns concerning faculty of study, in turn 3 the deceiver without marking the shift suddenly changed topics.

)5((Open new topics)

Deceiver:	เรียนคณะอะไร				
	เรียน|	คณะ|	อะไร|		
	riːan	kʰá ná	ʔà raj		
	study	faculty	what		
	‘What major do you study?’				First topic
Naïve partner:	วิทยา				
	วิทยา|				
	wít tʰá jaː				
	Sciences				
	‘Science’				
Deceiver:	แลกเบอร์กันนะ				
	แลก|	เบอร์|	กัน|	นะ|	
	lɛ̂ːk	bɤː	kan	Particle	
	Exchange number together				
	‘Let’s exchange (cell(numbers.				Second topic
Naïve partner:	ไม่บอก				
	ไม่|	บอก|			
	mâj	bɔ̀ːk			
	Not	tell			
	‘)I(won’t tell.’				

)6((Spontaneous correction)

Deceiver:	บอกชื่อเราย้วย			← Original typed text
Deceiver:	บอกชื่อเราด้วย			← A spontaneous correction
บอก|	ชื่อ|	เรา|	ด้วย|	
bɔ̀ːk	tɕʰɯ̂ː	raw	dûaj	
Tell	name	me	too	
‘Tell me (your(name too.’				
Naïve partner:	ชื่อพง			
	ชื่อ|	พง|		
	tɕʰɯ̂ː	pʰoŋ		
	Name	Pong		
	‘)My(name is Pong.			

In this example, the deceiver immediately corrected the last word from ย้วย /júaj/ ‘)to be(awry’)presumably a typo(to ด้วย /dûaj/ ‘too’.

Two linguistic features, demonstratives (e.g. นี้ /níː/ ‘this’(and self-reference frequency (e.g. ผม /pʰǒm/ ‘I’ used by men only), were also produced more frequently by deceivers than truth-tellers when they were engaging in online conversation. The participants as deceivers used demonstratives to provide a broad answer about where they were studying, as shown in Example (7). In Example 8, the deceiver used male self-reference to lie about her gender when chatting with the naïve partner.

)7((Demonstratives)

Naïve partner: เรียนที่ไหน

เรียน|	ที่ไหน|
ri:an	tʰîː năj
study	where

'Where do you study?'

Deceiver: เรียนนี่ ชลบุรี

เรียน|	นี่|	ชลบุรี|
ri:an	nîː	tɕʰon bùʔ riː
study	here	Chonburi

')I'm(studying here Chonburi'

)8((Frequency of Self-references)

ผมชื่อมอสครับ

ผม|	ชื่อ|	มอส|	ครับ|
pʰŏm	tɕʰûː	mɔ̂ːt	kʰráp
My	name	Mos	GP

'My name is Mos.'

Other results concern features with sub-categories where the chi-test was used to determine whether there were any significant differences between deceivers and truth-tellers in Thai online chat. Percentage differences were also used to examine which sub-categories of these features the deceivers used differently from the truth-tellers in Thai online chat. Table 3 presents the three features that show differences between deceivers and truth-tellers in Thai online chat. One is a linguistic feature derived from previous research into deception in English, and the other two are Thai-specific linguistic features.

Table 3: *Significant features with sub-categories used by deceivers and truth-tellers*

Difference in Linguistic Features									
24.	**Modal Verbs**	**Deceivers**	**Truth-tellers**	**Total**	***df***	**Chi-Square**	***P*-Value**	**Effect Size**	**Percentage Difference**
	High Confidence	25	17	42					47.05*
	Mid Confidence	51	85	136	2	11.92	.00*	.50	-40.00
	Low Confidence	17	8	25					112.50*
	Total	93	110	203					
Difference in Thai Specific-Features									
26.	**Self-Reference Types**	**Deceivers**	**Truth-tellers**	**Total**	***df***	**Chi-Square**	***P*-Value**	**Effect Size**	**Percentage Difference**
	1st Personal Pronouns	178	132	310					34.85*
	Kinship terms	34	12	46	3	16.73	.00*	.59	183.33*
	Using Nickname	72	74	146					-2.70
	Ellipsis Pronoun	561	584	1145					-3.94
	Total	845	802	1647					
27.	**Particles**	**Deceivers**	**Truth-tellers**	**Total**	***df***	**Chi-Square**	***P*-Value**	**Effect Size**	**Percentage Difference**
	Gender Particles	266	134	400					98.51*
	Info-oriented	78	71	149	3	37.53	.00*	.88	9.86
	General-oriented	385	417	802					-7.67
	Action-oriented	11	13	24					-15.38
	Total	740	635	1375					

Deceivers produced fewer mid-confidence modal verbs but more high-confidence and low-confidence modal verbs (see Examples 9 and 10), more self-reference especially for first-person pronouns and kinship terms (see Example 10), and a greater number of particles, especially gender particles (see Example 11), than truth-tellers in Thai online chat with statistically significant differences at the .01 level of significance. The effect size of these relationships are phi = .50, .59 and .88 respectively, all of which are considered statistically large effect sizes.

In (9), the deceiver used a modal verb with low confidence (อาจ /ʔàːt/ 'may') while chatting with a naïve partner.

)9((Modal verbs with low confidence)

พี่อาจหล่อกว่าที่คิดนะ

พี่|	อาจ|	หล่อ|	กว่า|	ที่|	คิด|	นะ|
pʰîː	ʔàːt	lɔ̀ː	kwâː	tʰîː	kʰít	náʔ
Older Sibling	may	handsome	more	that	think	particle

'I may be more handsome than what (you) think.'

In (10), the deceiver used a kinship term (พี่ /pʰîː/ 'older brother/sister') to pretend to be older than the naïve partner.

)10((Self-references types)

พี่พึ่งจบ

พี่|	พึ่ง|	จบ|
pʰîː	pʰɯ̂ŋ	tɕòp
Older sibling	just	graduate

'I have just graduated.'

In example (11), the deceiver used a male gender particle (GP) to lie about her actual gender while she was giving general information to the naïve partner.

)11((Particles)

เรียนเคมีครับ ลาดกระบัง

เรียน|	เคมี|	ครับ|	ลาดกระบัง|
riːan	kʰeːmîː	kráp	lâːt kràʔ baŋ
study	Chemistry	GP	Ladkrabang

')I(study Chemistry at Ladkrabang,'

In summary, the overall results show there were seven categories that appear to be associated with deception in Thai online chat. The results show that deceivers changed topics, made spontaneous correction, and used demonstratives, self-references, modal verbs with high and low confidence, self-reference types and gender particles more frequently than truth-tellers when engaging in Thai online chat.

9 Discussion

The objectives of this study were, firstly, to investigate whether the deceptive features found in English can be applied to identification deception in other languages, such as Thai, and, secondly, to find out whether there are any Thai-specific features that might be associated with deception in Thai online chat. Two comparable sets of deceivers' and truth-tellers' messages were collected while the participants were engaged in online conversation. We then categorized the results of the study examining deceptive indicators in Thai online chat into four major categories, depending on how they compare to previous research into deception detection.

First, findings in our study matching previous research into English include one communication behavior)opening new topics(and two linguistic features (demonstratives and modal verbs(. Possible explanations for these findings are that deceivers may open new topics frequently to dissociate themselves from the conversation (Twitchell & Nunamaker 2004), may use demonstratives frequently to avoid giving

specific information (Strömwall et al. 2004), and may use strong boosters to appear confident and strong hedges to protect their image (Buller & Burgoon 2006). These explanations appear to fit more closely to the Interpersonal Deception Theory (IDT) than to other theories of deception, suggesting that IDT may be more generalizable.

Second, findings showing no difference between deceivers and truth-tellers in contrast to previous research into English deception concern the majority of the investigated features. Our results indicate that, in Thai online chat, twenty features previously identified as associated with deception did not show any significant differences and could not differentiate deceivers and truth-tellers. These results may reflect a general lack of consistency of features identified in previous research into deception detection. In ten previous studies on English that examine at least twenty-five features, a specific feature is identified as associated with deception in only 28% of cases. This shows inconsistency of deceptive indicators in prior studies in English. The different tasks used in previous studies (e.g. desert survival experiments, giving false opinions, mock crimes and game simulations) may promote the use of different features, making task type as influential a variable whether a participant is being deceptive, and meaning that a definitive list of features associated with deception is unattainable.

Third, two findings standing in direct contrast to research into English are the frequency of self-reference and spontaneous correction. Our results show that self-reference was used by deceivers more frequently than by truth-tellers in Thai online chat. The traditional lie-detection research into English identified a decrease in deceivers' self-reference to avoid self-involvement when being deceptive in online conversation (e.g. Hancock et al. 2008). However, in our study, the high use of self-reference in Thai may be related to Thai-specific features because there are several types of self-reference in Thai. In English, self-reference is denotative with a purpose of associating or dissociating with the conversation (Wiseman 2007). In Thai, with several choices of self-reference, it is connotative reflecting gender as well as relative social relationship between chatters in online conversation. Our results, therefore, suggest that Thai online deceivers could use self-references frequently so as to manipulate their identity because they were pretending to be someone else in Thai online chat. The spontaneous correction produced by deceivers was also higher than truth-tellers, which is inconsistent with previous studies (e.g. Zhou & Zhang 2004). The reasons for this are unclear but could involve a desire to show the pretend identity accurately.

Finally, findings for Thai specific features associated with deception concern self-reference types and particles. The deceivers' use of first-person self-reference, kinship terms as self-references and gender particles have not been previously mentioned in studies of deception detection in English online communication. The Thai online deceivers in this study often used gender swapping, which may affect how they refer to themselves. For instance, first-person pronouns in Thai are mostly gender-specific, so their frequent use by deceivers could reflect a concern with creating an opposite gender identity. The same argument also applies to the frequent use of gender particles by the deceivers in this study. We then examined how often deceivers and truth-tellers used gender particles in Thai online chat. Table 4 shows that the more deceivers manipulated their gender identity in Thai online chat, the more they used gender particles as a deception strategy when pretending to be someone else engaging in online conversation.

Table 4: *Total number of gender particles as a proportion of all words used*

Participants	**No. of Gender Particles**
Truth-tellers	2.8%
Deceivers who did not change gender	5.0%
Deceivers who changed their gender	6.0%

Our results highlight the importance of not relying solely on analyses of English to investigate linguistic aspects of social behaviors. Deception is a multifaceted issue with the different facets of deception associated with different linguistic features. If a language does not have a feature that can manifest a certain facet of deception (as English does not have linguistic features for gender self-reference) then that facet will be overlooked in linguistic research on deception. This linking of features with facets does not apply solely to deception but could be applied to other issues such as identity construction, politeness and stance. We therefore believe that the current emphasis on using English as the main language for investigating issues in applied linguistics must be supplemented by investigations of the same issues in other languages.

10 Conclusion

In this paper, we have attempted to detect deception in Thai online chat. Some differences between language usage of deceivers and truth-tellers were found, but most of the features associated with deception in English online communication were not apparent in Thai deceptive language. We did, however, identify two Thai-specific linguistic features related to gender presentation which have not been previously reported. The use in Thai of features that identify the gender of the speakers' choice of self-reference and gender particles' allows deception researchers to analyze an aspect of deception, gender-swapping, which is not clearly marked in some other languages, such as English. These two features may be usefully applied in automated online deception detection for Thai and also highlight the need to supplement linguistic investigations of English with investigations of other typological distinct languages.

11 Acknowledgements

We would like to thank the Office of the Higher Education Commission, Thailand for supporting by grant fund under the program Strategic Scholarships for Frontier Research Network for the Ph.D. Program Thai Doctoral degree for this research.

References

Adkins, Mark, Douglas P. Twitchell, Judee K. Burgoon, & Jay F. Nunamaker, Jr. 2004. Advances in automated deception detection in text-based computer-mediated communication. In: Proceedings of the SPIE Defense and Security Symposium. August 13, 2004, Orlando, Florida.

Amos, Barrett. 2008. Are you lying now? A linguistic examination of deceptive utterances in online conversation. Honor's Thesis: Cornell University.

Aroonmanakun, Wirote. 2002. Thai word segmentation. Research Division of the Faculty of Arts, Chulalongkorn University, Thailand. Access http://pioneer.chula.ac.th/~awirote/resources/thai- word-segmentation.html

Buller, David B. & Judee K. Burgoon. 1996. Interpersonal deception theory. Communication Theory 6.3: 203-242.

Buller, David B. & Judee K. Burgoon. 2006. Interpersonal deception theory. In Griffin, Emory & Glen

Arthur McClish. *A first look at communication theory*. Boston: McGraw-Hill. 81-97.

Burgoon, Judee K. 2009. Interpersonal deception theory. In Littlejohn, Stephen W. & Karen A. Foss (eds.). *Encyclopedia of communication theory 1*. California: SAGE. 551-552.

Burgoon, Judee K. & TianTian Qin 2006. The dynamic nature of deceptive verbal communication. *Journal of Language and Social Psychology* 25.1: 76-96.

Burgoon, Judee K., Pete J. Blair, TianTian Qin & Jay F. Nunamaker, Jr. 2003. Detecting deception through linguistic analysis. In: *Proceedings of the Intelligence and Security Informatics*. June 2-3, 2003, Tucson, Arizona.

Burgoon, Judee K., David B. Buller & Kory Floyd. 2001. Does participation affect deception success? *Human Communication Research* 27.4: 503-534.

Celli, Fabio & Massimo Poesio. 2014 Pr2: A language independent unsupervised tool for personality recognition from text. arXiv preprint 1402.2796.1: 1-4

Charoenpornsawat, Paisarn. 2003. SWATH: Smart Word Analysis for Thai. Access http://www.cs.cmu.edu/~paisarn/software.html.

Cohen, Jacob. 1992. A power primer. Psychological Bulletin 112.1: 155-159

Cooke, Joseph R. 1989. Thai sentence particles: forms, meanings, and formal-semantic variation. Thai. Papers in Southeast Asian Linguistics, 12: 1-90

Dryer, Matthew S. & Haspelmath, Martin (eds.) 2013. The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. Access http://wals.info

Duran, Nicholas D., Charles Hall, Philip M. McCarthy & Danielle S. McNamara. 2010. The linguistic correlates of conversational deception: Comparing natural language processing technologies. Applied Psycholinguistics, 31.3: 439-462.

Ekman, Paul & Wallace V. Friesen. 1969. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. Semiotica 1.1: 49-98.

Ekman, Paul & Wallace V. Friesen. 1974. Detecting deception from the body or face. Journal of Personality and Social Psychology 29.3: 288-298.

Fornaciari, Tommaso & Massimo Poesio. 2011. Sincere and deceptive statements in Italian criminal proceedings. In: Proceedings of the International Association of Forensic Linguists Tenth Biennial Conference, IAFL. July 11-14, 2011, Aston University, Birmingham.

Hancock, Jeffrey T., Lauren E. Curry, Saurabh Goorha & Michael T. Woodworth. 2005. Automated linguistic analysis of deceptive and truthful synchronous computer-mediated communication. In: Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICCS'05). January 3-6, 2005, Hawaii.

Hancock, Jeffrey T., Lauren E. Curry, Saurabh Goorha & Michael T. Woodworth. 2008. On lying and being lied to: A linguistic analysis of deception in computer-mediate communication. Discourse Processes 45.1: 1-23.

Hargitai, Rita, Mátyás Naszódi, Balázs Kis, László Nagy, Adrienn Bóna & János László. 2007. Linguistic markers of depressive dynamics in self-narratives: Negation and self-reference. Empirical Text and Culture Research 3:26-38.

Hauch, Valerie, Jaume Masip, Iris Blandon-Gitlin & Siegfried L. Sporer 2012. Linguistic cues to deception assessed by computer programs: a meta-analysis. In: Proceedings of the workshop on computational approaches to deception detection.

Hoonchamlong, Yuphaphann. 2005. E-Thai: Thai Language in the new millennium. In: Proceedings of the 9th International Conference on Thai Studies. April 3-6, 2005, Illinois.

Internet Information Research. 2015. Internet users and statistics in Thailand. Access http://internet.nectec. or.th/webstats/home.iir

Iwasaki, Shoichi & Preeya Ingkaphirom. 2005. A reference grammar of Thai. Cambridge: Cambridge University Press.

Johnson, Deborah G. 1997. Ethics online. Communications of the ACM 40.1: 60-65.

Kassin, Saul M. & Christina T. Fong. 1999. "I'm innocent!" Effects of training on judgments of truth and deception in the interrogation room. Law and Human Behavior 23.5: 499-516.

Köhnken, Günter. 2004. Statement validity analysis and the detection of the truth. In Grandhag, Pär Andre & Leif A. Strömwall (eds.(. The detection of deception in forensic contexts. Cambridge: Cambridge University Press. 41-63.

Larson-Hall, Jenifer.)2012). How to run statistical analyses. In Mackey Alison and Susan M. Gass (eds.(. Research methods in second language acquisition. Oxford: Wiley-Blackwell. 245-274.

National Electronics and Computer Technology Center (NECTEC). 2010. Internet user profile of Thailand 2009. Prathumtani, Thailand.

Newman, Matthew L., James W. Pennebaker, Diane S. Berry & Jane M. Richards. 2003. Lying words: Predicting deception from linguistic styles. Personality and Social Psychology Bulletin 29.5: 665-675.

Ochs, Elinor & Bambi Schieffelin. 1989. Language has a heart. Text-Interdisciplinary Journal for the Study of Discourse 9.1: 7-26.

Panyametheekul, Siriporn. 2005. Patterns of Thai greetings in Thai chat room. Paper presented at Language, Society, and Communication: Pragmatics, Sociolinguistics and Discourse Studies. July 29-30, 2004, Chulalongkorn University, Bangkok, Thailand.

Pennebaker, James W., Tracy J. Mayne & Martha E. Francis. 1997. Linguistic predictors of adaptive bereavement. Journal of Personality and Social Psychology 72.4: 863-871.

Philipson, Robert. 1992. Linguistic imperialism. Oxford: Oxford University Press.

Rabon, Don. 2003. Investigative discourse analysis. North Carolina: Carolina Academic Press.

Reed, Alan. 2012. Simple Concordance Program (Version 4.09). Access http://www.textworld.eu/scp/index.html

Richards, Jane M. & James J. Gross. 1999. Composure at any cost? The cognitive consequences of emotion suppression. Personality and Social Psychology Bulletin 25.8: 1033-1044.

Rubin, Victoria L. 2014. Pragmatic and cultural considerations for deception detection in Asian Languages. ACM Transactions on Asian Language Information Processing 13.2:1-10.

Rungruangthum, Montarat & Richard Watson Todd. 2011. Cyber deception: Linguistic analysis of truthful and deceptive messages in Thai online chat. In: Proceedings of the Joint International Symposium on Natural Language Processing and Agricultural Ontology Service, SNLP-AOS. December 7-8, King Mongkut's University of Technology Ladgrabang, Thailand.

Rungruangthum, Montarat, Richard Watson Todd & Wirote Aroonmanakun. 2011. Re-conceptualization of piloting research in applied linguistics. In: Proceedings of Doing Research in Applied Linguistic, DRAL. April 21-22, King Mongkut's University of Technology Thonburi, Thailand.

Sinclair, John. 1991. Corpus, concordance, collocation. Oxford: Oxford University Press.

Sip, Kamila E., Andreas Roepstorff, William McGregor & Chris D. Frith. 2008. Detecting deception: The scope and limits. Trends in Cognitive Sciences 12.2: 48-53.

Smyth, David. 2002. Thai: An essential grammar. London: Routledge.

Sornlertlamvanich, Virach, Naoto Takahashi & Hitoshi Isahara. 1997. ORCHID: Thai part-of-speech tagged corpus. TR-NECTEC 1997001: 5-19.

Sornlertlamvanich, Virach, Naoto Takahashi & Hitoshi Isahara. 1999. Building a Thai part-of-speech tagged corpus (ORCHID) Journal of the Acoustical Society of Japan 20.3: 189-198.

Spence, Katelyn, Gina Villar & Joanne Arciuli. 2012. Markers of deception in Italian speech. Frontiers in Psychology 3:1-9.

Strömwall, Leif A., Pär Andre Granhag & Maria Hartwig. 2004. Practitioners' beliefs about deception. In Grandhag, Pär Andre & Leif A. Strömwall (eds.(. The detection of deception in forensic contexts. Cambridge: Cambridge University Press. 229-250.

Thammawan, Suchada. 2008. A study of the differences of some grammatical structures of English and Thai that cause English speaking problems for Thai people. Master's Thesis: Thammasat University, Thailand.

Tesprasit, Virongrong, Paisarn Charoenpornsawat & Virach Sornlertlamvanich, 2003. Learning phrase break detection in Thai text-to-speech. Paper presented at the INTERSPEECH. September 1-4, 2003, Geneva, Switzerland.

Thailand Technology Crime Suppression Division. 2011. Internet crime statistics in 2011. National report: Bangkok, Thailand.

Twitchell, Douglas P., Jay F. Nunamaker, Jr. & Judee K. Burgoon. 2004. Using speech act profiling for deception detection. Paper presented at the International Conference on Intelligence and Security Informatics. June 10-11, 2004, Arizona.

Twitchell, Douglas P. & Jay F. Nunamaker, Jr. 2004. Speech act profiling: a probabilistic method for analyzing persistent conversations and their participants. In: Proceedings of the 37th Annual Hawaii International Conference on System Sciences. January 7-10, 2004, Hawaii.

Vrij, Aldert, Samantha Mann, Susanne Kristen & Ronald P. Fisher. 2007. Cues to deception and ability to detect lies as a function of police interview styles. Law and Human Behavior 31.5: 499-518.

Vrij, Aldert, Ronald Fisher, Samantha Mann, & Sharon Leal. 2008. A cognitive load approach to lie detection. Journal of Investigative Psychology and Offender Profiling 5.1-2: 39-43.

Wiseman, Richard. 2007. Quirkology: the curious science of everyday lives. London: Macmillan.

Yates, Dan, David S. Moore & Daren S. Starnes 1999. The practice of statistics. New York: Freeman.

Yimngam, Sukanya., Wichian Premchaisawadi & Worapoj Kreesuradej. 2009. Thai emotion words analysis. In: Proceedings of the 8th International Symposium on the Natural Language Processing (SNLP'09). October 20-22, 2009, Dhurakij Pundit University, Bangkok, Thailand.

Zhou, Lina. 2005. An empirical investigation of deception behavior in instant messaging. IEEE Transactions on Professional Communication 48.2: 147-160.

Zhou, Lina, Judee K. Burgoon, Jay F. Nunamaker, Jr. & Douglas P. Twitchell. 2004a. Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications. Group Decision and Negotiation 13.1: 81-106.

Zhou, Lina, Judee K. Burgoon, Douglas P. Twitchell, Tiantian Qin & Jay F. Nunamaker, Jr. 2004b. A comparison of classification methods for predicting deception in computer-mediated communication. Journal of Management Information Systems 20.4: 139-166.

Zhou, Lina & Yu-wei Sung. 2008. Cues to deception in online Chinese groups. In: Proceedings of the 41st Hawaii International Conference on System Sciences. January 7-10, 2008, Hawaii.

Zhou, Lina, Douglas P. Twitchell, Tiantian Qin, Judee K. Burgoon & Jay F. Nunamaker, Jr. 2003. An exploratory study into deception detection in text-based computer-mediated communication. In: Proceedings of the 36th Annual Hawaii International Conference on System Sciences. January 6-9, 2003, Hawaii.

Zhou, Lina & Dongsong Zhang. 2004. Can online behavior unveil deceivers? An exploratory investigation of deception in instant messaging. In: Proceedings of the 37th Annual Hawaii International Conference on System Sciences. January 5-8, 2004, Hawaii.

Zhou, Lina & Dongsong Zhang. 2008. Following linguistic footprints: Automatic deception detection in online communication. Communications of the ACM 51.9: 119-122.

Zuckerman, Miron, Bella M. DePaulo & Robert Rosenthal. 1981. Verbal and nonverbal communication of deception. Advances in Experimental Social Psychology 14.1: 1-59.

Reviewed: Received 11 September 2017, revised text accepted 11 November 2017, published 10 December 2017
Editors: Editor-In-Chief Dr Mark Alves | Managing Eds. Dr Paul Sidwell, Dr Nathan Hill, Dr Sigrid Lew